Current Activities in the Scheduling and Resource Management Area of the Global Grid Forum

Bill Nitzberg[1] and Jennifer M. Schopf[2]

[1] Veridian, PBS Products, 2672 Bayshore Pkwy, Suite 810, Mountain View, CA 94043, USA
bill@computer.org
[2] Mathematics and Computer Science Division, Argonne National Laboratory, 9700 S. Cass Ave., Argonne, IL 60439, USA
jms@mcs.anl.gov

Abstract. The Global Grid Forum's Scheduling and Resource Management Area is actively pursuing the standards that are needed for interoperability of Grid resource management systems. This includes work in defining architectures, language standards, APIs and protocols. In this article we overview the state of the working groups and research groups in the area as of September 2002.

1 Introduction

The Global Grid Forum (GGF) [1] is an open standards body focused on Grid computing. Organized similarly to the Internet Engineering Task Force (IETF) [2], the GGF consists of groups of committed individuals from academia, research labs, and industry, working toward standards to promote common understanding and, more importantly, interoperability. The current areas in the GGF are Architecture, Data, Information Systems and Performance, Peer-to-peer Computing, Scheduling and Resource Management, Security, and Applications, Programming Models and Environments.

The main focus of the Scheduling and Resource Management Area is agreements and standards in Grid resource management: architecture, specifications for resources and requirements, queuing, scheduling and superscheduling, starting and stopping jobs (task management), and accounting and logging. Generally, the process begins by looking at what is done today and what is desired; then gathering requirements, refining protocols, interactions, capabilities, and the like, and, finally, working to standardize APIs and protocols. Overall, the goal of this area is to "enable better use of resources". The current makeup of active participants in the Scheduling Area covers Grid "operating systems" level developers, researchers, application developers, students, Grid system managers, and a smattering of others. All GGF activities are open; anyone is welcome to participate (visit www.gridforum.org).

The "output" of Global Grid Forum activities is documents relating to Grid standards.

It is important to understand how different levels of standardization promote interoperability: models (or frameworks, architectures) create a human-level common understanding among people, APIs (or interfaces) enable code portability and re-use, but not necessarily interoperability between different code bases; protocols enable interoperability between different code bases, but not necessarily code portability; and, languages (or sets of tokens) are a building block for all of the above.

For instance, MPI [3] is an example of a standard API—allowing a programmer to write a single parallel program that can run either on a cluster of Linux machines with MPICH or on an IBM SP with IBM's proprietary MPI implementation. A simple recompilation is all that is required. MPI does not, however, support communication between two MPI programs (one on each of the above systems); that is, it supports code portability, not interoperability. TCP/IP, on the other hand, is a standard protocol that supports interoperability but says nothing about code portability. A program running on Microsoft Windows using the WinSock API can easily communicate with another program running on UNIX that uses the sockets API.

2 Current GGF Scheduling and Resource Management Area Efforts

As of GGF-5, July 2002, the Scheduling Area had two finishing groups, four active groups and five new groups proposed, with over two hundred people participating. Roughly, these activities fall into the following categories:

Architecture and Overview
- "Ten Actions When Superscheduling" (<u>Architecture</u>, completed Group)
- Scheduling Dictionary Working Group (<u>Language</u>)
- Grid Scheduling (<u>Architecture</u>, proposed Group)
- Grid Economic Brokering (<u>Architecture</u>, proposed Group)

Standards to "run my job"
- Advance Reservation API Working Group (<u>API</u>, completed Group)
- Distributed Resource Management Application API Working Group (<u>API</u>)
- Grid Resource Allocation and Agreement Protocol Working Group (<u>Protocol</u>)

Super-Scheduling
- Scheduling Attributes Working Group (<u>Language</u>)
- Scheduling Optimization (proposed Group)

Basic accounting for interoperability
- Usage Record (<u>Language</u>, proposed Group)
- OGSA Resource Usage Service (<u>Protocol</u>, proposed Group)

The first document prepared by the Scheduling Area was the "Ten Actions When Superscheduling" document [4], led by J. Schopf. This document outlines the steps a user goes through when scheduling across Grid resources; the basic steps are shown in Figure 1. These are grouped into three phases – resource discovery, system selection, and running a job – and spell out the basic steps of scheduling a job. This document is in final review and has been updated and extended for publication in a journal special issue as well [5].
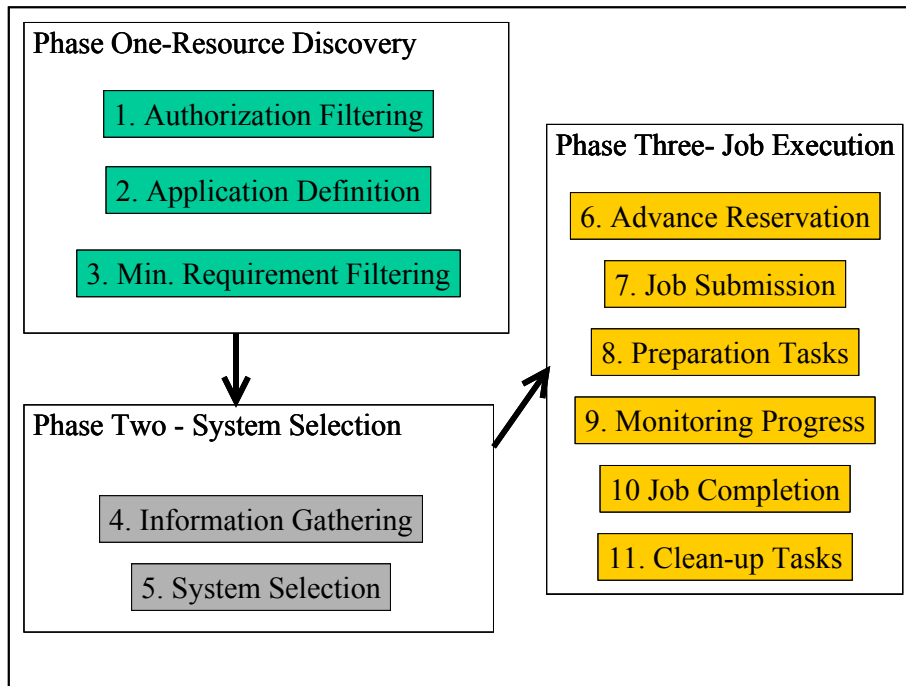
Phase One-Resource Discovery

1. Authorization Filtering

2. Application Definition

3. Min. Requirement Filtering

Phase Two - System Selection

4. Information Gathering

5. System Selection

Phase Three- Job Execution

6. Advance Reservation

7. Job Submission

8. Preparation Tasks

9. Monitoring Progress

10 Job Completion

11. Clean-up Tasks

Fig. 1. Ten Steps for Superscheduling.

Another document that is in the final review process is the "Advance Reservation API", by A. Roy and V. Sander [6]. This document defines an experimental API for quality of service reservations for different types of resources. It is strongly based on GARA [7]. This document is in final stages of review.

The active working group, Scheduling Attributes [8], led by U. Schwiegelshohn and R. Yahyapour, is defining a set of attributes of lower-level scheduling instances that can be used to make resource management decisions by higher-level schedulers. The document created by this group [9] in final stages of review.

The Scheduling Dictionary working group [10], led by Wieder and Ziegler, is identifying and defining the terms needed to discuss schedulers. Early on, we observed that each researcher in the area used the same terms in slightly different ways.

The goal of this group is to aid interoperability (especially among people working in this field). This group has a draft of it's document available online.

The Distributed Resource Management Application API group (DRMAA) [11], led by J. Tollefsrud and H. Rajic, is defining an API for the submission and control of jobs to one or more distributed resource management systems. They plan to present a semi-final draft presented at GGF-6 in Chicago.

J. Maclaren, V. Sander, and W. Ziegler lead the working group on Grid Resource Allocation Agreement Protocol [12]. This group is defining the interactions between a higher-level service and a local resource management system. The goal is to facilitate the allocation and reservation of Grid resources. Much of this work is growing out of the SNAP [13] work as well.

At GGF-5 in Scotland, 5 groups were proposed as part of the Scheduling Area. U. Schwiegelshohn proposed the Grid Scheduling Architecture working group [14]. This group will define an architecture that details the interactions between a Grid scheduler and other components, such as a Grid information system, a local resource management system, and network management systems. This group is awaiting full development of a charter and assessment of interest.

Three groups related to accounting issues were proposed. The first, and cornerstone to the others, is the Usage Record working group [15], presented by L. McGinnis. The goal of this group is to define a common accounting usage record (format and contents) to promote the exchange of accounting information between sites. This isn't to replace the records that are being used at current sites, but is to be used to exchange them. The TeraGrid project [16] has identified this as a key need.

A second group related to accounting issues is the proposed Grid Economic Service Architecture working group [17], currently being led by S. Newhouse, J. MacLaren, and K. Keahey. This architecture-focused group will define a supporting infrastructure that enables organizations to "trade" services between each other. The infrastructure will include the definition of protocols and service interfaces that will enable the exploration of different economic mechanisms (but not the economic models). A charter for this group is being finalized.

The third accounting-focused group is the OGSA Resource Usage Service [18], with proposed chairs of S. Newhouse and J. Magowan. To track resource use within OGSA Grid services, we need to develop a service interface that supports the recording and retrieval of resource usage. The charter for this group is being finalized.

The fifth group proposed was a research group on the topic of Scheduling Optimisation [19], led by V. Di Martino and E. Talbi. This group proposes to define measures of scheduling algorithm performance and to foster the development of Grid-wide scheduling methodology on top of available schedulers.

3 Fruitful Directions – What's Next?

We expect the UR and DRMAA activities to complete this year and have a positive impact on the community. The ability to exchange Usage Record (accounting) data between sites participating in Grid activities is a fundamental prerequisite to achieving acceptance and commitment of resources from both the funding agencies and the resource owners. The proposed UR group already has active participation from the TeraGrid, NASA's IPG, and industry. DRMAA will greatly ease the burden on the applications programmer's use of resource management systems and will foster third-party Grid-enabled commercial products. DRMAA has strong industry participation (including Sun, Intel, Veridian, Cadence, and HP).

Looking outside the current activities within the GGF, we believe the following would be fruitful directions:

- Language for resource and job specification – many different languages exist today; a standard language to promote interchange between existing systems would enable easier job migration among these distinct systems.
- API for scheduling (especially for superscheduler-scheduler interaction) – not only would this ease implementation of superschedulers, but it would also enable "research" schedulers to be plugged into production environments for real-world experience.
- Language for describing site-specific scheduling policy and requirements – tuning any scheduling system is a complicated, iterative process; a standard language would allow one to duplicate policies at different sites, each using its own resource management system, and, in the longer term, would allow a superscheduler to reason about site policies.
- Agreements on resource fungability – to enable economy-based trading of resources. (The proposed GESA working group may attack this topic.)
- Work on Grid-level policy management across scheduling systems.

The best standards build on existing work. Over the next ten years, we expect a snowball effect as the work coming out of the Global Grid Forum excites the community to explore new directions.

4 How to Become Involved

GGF participants come from over two hundred organizations in over thirty countries, with financial and in-kind support coming from GGF Sponsor Members including commercial vendors and user organizations as well as academic and federal research institutions. Anyone interested in Grid computing, or in the Global Grid Forum activities specifically, is welcome to participate in a GGF meeting or event.

To join the GGF Scheduling and Resource Management Area mailing list, please send mail to majordomo@gridforum.org with the message "subscribe sched-wg".

Acknowledgments

References

[1] The Global Grid Forum, www.gridforum.org

[2] The Internet Engineering Task Force, www.ietf.org

[3] Message Passing Interface Forum, MPI-2: Extensions to the Message-Passing Interface, Sept. 2001, http://www.mpi-forum.org/docs/mpi-20-html/mpi2-report.html

[4] Schopf, J. M., "Super Scheduler Steps/Framework", Global Grid Forum Scheduling Area Working Document SchedWD 8.5, http://www-unix.mcs.anl.gov/~schopf/ggf-sched/WD/schedwd.8.5.pdf

[5] Schopf, J. M., "A General Architecture for Scheduling on the Grid", submitted to special issue of the Journal of Parallel Distributed Computing on Grid Computing, 2002. Available from http://www-unix.mcs.anl.gov/~schopf/Pubs/sched.arch.jpdc.pdf

[6] Roy, A., and Sander, V., "Advanced Reservation API", Scheduling Area Working Document SchedWD 9.4, http://www-unix.mcs.anl.gov/~schopf/ggf-sched/WD/schedwd.9.4.pdf

[7] Roy, Alain, "End-to-End Quality of Service for High-End Applications", Ph.D. dissertation, University of Chicago, Department of Computer Science, August 2001.

[8] Global Grid Forum, Scheduling Attributes Working Group, Scheduling and Resource Management Area, http://ds.e-technik.uni-dortmund.de/~yahya/ggf-sched/WG/sa-wg.html

[9] Schwiegelshohn, U. and Yahyapour, R., "Attributes for Communication about Scheduling Instances", Scheduling Area Working Document SchedWD 10.5, http://www-unix.mcs.anl.gov/~schopf/ggf-sched/WD/schedwd.10.5.pdf

[10] Global Grid Forum, Scheduling Dictionary Working Group, Scheduling and Resource Management Area, http://www.mcs.anl.gov/~jms/ggf-sched/WG/sd-wg.html

[11] Global Grid Forum, Distributed Resource Management Application API Working Group, Scheduling and Resource Management Area, http://www.mcs.anl.gov/~jms/ggf-sched/WG/drmaa-wg.html

[12] Global Grid Forum, Grid Resource Allocation Agreement Protocol Working Group, Scheduling and Resource Management Area, http://people.man.ac.uk/~zzcgujm/GGF/graap-wg.html

[13] Czajkowski, K., Foster, I., Kesselman, C., Sander, V., and Tuecke, S., "SNAP: A Protocol for Negotiating Service Level Agreements and Coordinating Resource Management in Distributed Systems", 8th Workshop on Job Scheduling Strategies for Parallel Processing, Edinburgh, Scotland, July 2002.

[14] Global Grid Forum, Proposed Grid Scheduling Architecture Working Group, Scheduling and Resource Management Area, http://ds.e-technik.uni-dortmund.de/~yahya/ggf-sched/WG/arch-rg.html

[15] Global Grid Forum, Proposed Usage record Working Group, Scheduling and Resource Management Area, http://www.mcs.anl.gov/~jms/ggf-sched/WG/ur-wg.html

[16] The TeraGrid Project, www.teragrid.org

[17] Global Grid Forum, Proposed Grid Economic Brokering Architecture Working Group, Scheduling and Resource Management Area, http://www.mcs.anl.gov/~jms/ggf-sched/WG/geba-wg.html

[18] Global Grid Forum, Proposed OGSA Resource Usage Service Working Group, Scheduling and Resource Management Area, http://www.mcs.anl.gov/~jms/ggf-sched/WG/rus-wg.html

[19] Global Grid Forum, Proposed Scheduler Optimization Research Group, Scheduling and Resource Management Area, http://www.mcs.anl.gov/~jms/ggf-sched/WG/opt-wg.html